

# 基于 copulas 技术的中国沪深股市 相关性分析\*

田 萍 张屹山

(吉林大学商学院；吉林大学数量经济研究中心)

## 一、引言

由于金融市场产品的复杂性和计算水平的提高，使人们越来越重视对各种的金融产品之间整体关系的研究。因为对这种关系的正确研究不但可以优化投资组合更可以使投资者很好的规避风险。当然，对于科研人员与市场管理者来说，正确地了解市场的整体运作形式能够有利于对市场的正确管理和保证经济市场平稳合理的运行。常见的一些研究主要包括如对市场的羊群效应研究、领滞关系研究以及整体风险和有效性研究等。国内外关于这方面的文献也数不胜数。长时间以来尽管随机变量的分布理论在处理单种资产收益率行为等的研究中被广泛应用，并取得了不错的成绩。但是在多种资产的市场研究时很少应用联合分布的条件进行分析，这并不是因为联合分布不符合实际经济关系的假设，而主要是由于多元分布本身的复杂性和不可知性决定的。因此，即便研究者偶尔应用多元联合分布的假设，也通常是二元正态的情况，这也限制了其边际分布必须是正态，而且是线性的（尽管这种假设并非十分合理）。但是，随着经济社会复杂情况的加剧以及计算机水平的日益提高，使人们想要获得复杂的多元变量之间准确的相互关系的要求更加强烈了，这不仅包含变量间的线性关系，还有非线形性和尾部相关性等，所有这些问题已不仅仅是正态假设及线性相关系数能够应付得了的了。

\* 本文部分受社科项目《前瞻性货币政策规则在我国的适应性研究》项目，基金号 07BJY168、中国博士后基金二等项目、吉林大学“985 工程”建设项目以及“经济分析与预测哲学社会科学创新基地”项目资助。

目前国际上正在发展的 copulas 函数技术对解决多个变量间复杂的相关关系问题有很大的帮助，它通过利用边际分布和已知的 copulas 函数进行多元关系研究，克服了直接研究多元分布函数的弊端。它不仅能够处理边际分布相同的多元分布问题，甚至对边际分布不同的多元联合分布问题也能够处理。又由于 copulas 函数的特殊性质，使人们对除线性关系以外的其他形式的变量间的相依关系的研究也有很大的帮助。因此，近些年来，关于 copulas 函数技术在金融的风险度量<sup>[4]、[7]、[11]、[13]</sup>、衍生工具定价<sup>[2]</sup>、生存模型<sup>[3]</sup>及保险<sup>[3]、[5]</sup>等方面都有很多重要且有意义的研究。可以预见 copulas 函数在金融市场的研究工作中将会起到越来越重要的作用。我国也有很多学者已开始重视 copulas 函数并关注其在金融方面的应用<sup>[11]~[15]</sup>。

本文在对 copulas 函数极其性质进行简单介绍的基础上，对中国的沪、深股市指数的相互关系进行参数估计，以便获得与其对应的几种可能的 copulas 函数形式，并就所得结果进行了简单的分析。

## 二、copulas 简介

简单地说，一个 copulas 函数实质上是一个连接函数，其把一个多元分布函数和与之对应的一维边际分布函数连接到一起。copula 一词最早于 1959 年在 Sklar, A 的理论中被提出。而 copulas 之所以能够被广泛的研究和应用的主要原因正如 Fisher (1997) 在《Encyclopedia of Statistical》<sup>[8]</sup> 中所说的“copulas 函数有两点主要原因使其获得统计学家的关注：‘一、它是研究自由尺度相关性的一种方法；二、它是研究变量分布族结构的出发点。’”

### (一) 模型介绍 (以二元随机变量为例)

设  $X, Y$  是二元随机变量分别具有分布函数  $F(x) = P(X \leq x), G(y) = P(Y \leq y)$  和联合分布函数  $H(x, y) = P(X \leq x, Y \leq y), \forall (x, y) \in (-\infty, \infty)^2$ 。下面考察  $(F(x), G(y), H(x, y))$  所处的  $I^3$  ( $I = [0, 1]$ ) 空间：称从  $I^2$  中把  $(F(x), G(y))$  映射成  $I$  中  $H(x, y)$  的函数为一个 copulas，或称为一个连接函数。公式表示为  $H(x, y) = C(F(x), G(y))$  或者写成  $C(u_1, u_2) = H(F^{-1}(u_1), G^{-1}(u_2)), \forall (u_1, u_2) \in [0, 1]^2$ 。其中， $F^{-1}, G^{-1}$  形如  $F^{-1}(p) = \sup\{x | F(x) \leq p\}$ 。由 Schweizer & Sklar (1983) 给出的定理证明了当  $F, G$  是连续函数时 Copula 函数唯一；当  $F, G$  不连续时，这种 copulas 表示不唯一。copulas 函数的一个优点是其不但能够表示变量之间的相关性，而且这种体现不会因为经过对变

量的严格的单调变换而改变，这是简单的线形相关系数做不到的。

## (二) 几个有意义的或本文的实证研究将会用到的 copulas 函数：

### 1. 独立分布的 copulas:

$$C_{\perp}(u_1, u_2) = u_1 u_2 \quad (1)$$

当随机变量相互独立时其 copulas 函数满足公式 (1)。

### 2. 完美相依的 copulas——Fréchet Copulas:

$$C_l(u_1, u_2) = \max \{ (u_1 + u_2 - 1), 0 \} \quad (2)$$

$$C_u(u_1, u_2) = \min(u_1, u_2) \quad (3)$$

简单地说，其中公式 (2) 体现了变量间完美的逆向相关的关系，而公式 (3) 却体现了变量间的完美的同向相关的关系，这两个 copulas 函数也正是所有的 copulas 函数的下、上界。因此，鉴于它们与独立分布的 copulas 函数的特殊性，虽然本文关心的沪、深指数的关系不适合使用，我们也在里给出简单的介绍。

### 3. Gaussian copulas:

$$C_G(u_1 u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp \left\{ -\frac{(s^2 - 2\rho st + t^2)}{2(1-\rho^2)} \right\} ds dt \quad (4)$$

其中， $-1 < \rho < 1$  表示两变量之间的相关性，但由于这个 copulas 函数已不单单表示具有线性关系的变量之间的联合分布形式了，因此参数  $\rho$  也不再是两个变量之间的线性相关系数。而  $\Phi(\cdot)$  为标准正态随机变量的分布函数。

### 4. t 分布 copulas:

$$C_t(v)(u_1, u_2) = \int_{-\infty}^{t_{v1}^{-1}(u_1)} \int_{-\infty}^{t_{v2}^{-1}(u_2)} \frac{(2\pi)^{-1}}{(1-\rho^2)^{1/2}} \exp \left\{ 1 - \frac{(s^2 - 2\rho st + t^2)}{v(1-\rho^2)} \right\}^{-\frac{v+2}{2}} ds dt \quad (5)$$

其中，参数  $\rho$  同公式 (4)， $t_{v1}^{-1}$  表示自由度为  $v_1$  的分位数。而  $v$  也表示自由度，待估。

### 5. Gumbel copulas:

$$C_G(u_1, u_2) = \exp \{ - [\tilde{u}_1^\delta + \tilde{u}_2^\delta]^{1/\delta} \} \quad (6)$$

其中， $\tilde{u} = -\log(u)$ 。 $\delta \geq 1$  体现了变量之间的相依性：当  $\delta=1$  时，两个变量相互独立。 $\delta \rightarrow \infty$  时表示变量是完美相依的。Gumbel 是一个极值 copulas，这种函数能够表现在紧张的市场状态（“牛市”或“熊市”）中资产的尾部相关性。该 Gumbel copulas 表现的是上尾相关，相关程度为  $2 - 2^{1/\delta}$ 。

### 三、copulas 函数的相关性度量

前面我们已经提到了 copulas 强于线性函数的地方是其能够正确的得到即便对变量进行了严格的单调变换后的变量的相关性，这无疑扩大了人们研究的范围。当然除了 copulas 函数外还有 Kendall 的  $\tau$ 、Spearman 的  $\rho_s$  和基尼系数  $r$  也都是单调变换不变的相关性度量指标，并且可以证明它们分别与 Copula 函数满足如下的关系：

$$\tau = 4 \iint_{\substack{0 \leq u_1, u_2 \leq 1 \\ u_1 + u_2 \geq 1}} C(u_1, u_2) dC(u_1, u_2) - 1 \quad (7)$$

$$\rho_s = 12 \iint_{\substack{0 \leq u_1, u_2 \leq 1}} u_1 u_2 dC(u_1, u_2) - 3 \quad (8)$$

$$r = 4 \left\{ \int_0^1 C(u, 1-u) du - \int_0^1 (u - C(u, u)) du \right\} \quad (9)$$

显然，有了上面的指标间的关系表达式无论对验证或选择合适的 copulas 函数都有很大的帮助。而除了对相关系数的表述特征外，某些 copulas 函数还有能够表现变量间尾部相关性的能力。可以说尾部相关性在金融市场的分析中是非常有意义的一个指标，用条件概率  $P(Y > y | X > x)$  或  $P(Y < y | X < x)$  表示，若在股票市场中它就反映了一种股票的价格大幅上涨或下跌后，是否会引起其他股票价格的攀升或下跌。当  $x, y$  相当大或相当小时表示的就是随机变量的尾部相关性，用  $\lambda(u)$  与  $\lambda(l)$  分别表示概率  $p(Y > u | X > u)$  和  $p(Y < l | X < l)$ ，则  $u \rightarrow \infty$  和  $l \rightarrow -\infty$  时  $\lambda(u)$  和  $\lambda(l)$  的极限值如果存在的话就反映了尾部相关性的大小，因此用：

$$\lambda_u = \lim_{u \rightarrow \infty} \lambda(u) \quad (10)$$

$$\lambda_L = \lim_{L \rightarrow \infty} \lambda(L) \quad (11)$$

表示尾部相关性。也可以证明  $\lambda(u)$  和  $\lambda(l)$  的极限是严格单调增变换不变的，因此它的极限满足连接函数的性质，可以用连接函数的极限值来表示。可以证明当极限存在时有：

$$\lambda_u = \lim_{\alpha \rightarrow 0^+} P(Y > G^{-1}(1-\alpha) \mid (X > F^{-1}(1-\alpha))) = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1-u} \quad (12)$$

$$\lambda_L = \lim_{\alpha \rightarrow 0^+} P(Y < G^{-1}(\alpha) \mid (X < F^{-1}(\alpha))) = \lim_{u \rightarrow 0} \frac{C(u, u)}{u} \quad (13)$$

前面我们介绍的 t 分布 copulas 和 Gumbel copulas 都是具有尾部相关性的

copulas 函数，并且其相关性程度会因参数值的不同而不同。通过对尾部相关性的引入就可以明显的区分出正态分布在处理股票市场联合分布时的弱点来，这是因为如果用  $\rho_{X,Y}$  表示随机变量  $X, Y$  的相关系数，不论这个相关系数  $\rho_{X,Y}$  有多大，只要不等于 1，它的  $\lambda_u$  和  $\lambda_l$  就等于 0。而对于非正态分布的厚尾分布如 t 分布而言情况就不一样了，它会根据自变量的相关系数的不同有不同的尾部相关性。

#### 四、中国股市沪、深市场指数相关性的 copulas 实证研究

本文开始就已经提到了 copulas 在金融市场的投资、管理等很多方面都可以进行应用，并且也已经取得了很不错的成绩，甚至一些公司已经开发这方面的软件，正如张尧庭老师所讲的它很快应会变成一种实用的技术。然而，对 copulas 的正确应用的首要条件就是能够获得准确符合实际变量的 copulas 函数，因此对应用者来说进行合理的估计以得到准确的 copulas 函数的参数也是非常重要的。因此，本文根据沪、深股市的具体特点以及通常对市场的假设条件的基础上探讨它们所服从的 copulas 函数形式并进行参数估计和简单的分析。文中所涉及算法的部分均用 Eviews 和 MatLab 软件完成。

##### (一) 简单的数据分析

本文所选的数据分别为上证综合指数和深证成分指数的从 1996 年 1 月 2 日～2001 年年底的每天的收盘数据共 1454 个。我们首先对普通的指数数据进行变化得到两个市场的每天的对数收益率数据 1453 个。首先我们可以得到收益率的散点图 1。

通过图 1 我们可以直观的看出在中国股票市场上的沪、深两个市场的收益率是存在明显的正相关关系的。使用样本数据获得两个市场对数收益率的线性相关系数为 0.845，这说明这两个市场的收益率存在程度很强的正相关性。但是正如前面所提到的线性相关系数仅能体现被考察者直观的共同变化趋势却不稳定，并且不能够表现如尾部相关性等特征。后面的进一步分析也能够体现这一特点。

##### (二) 相关于 copulas 函数的分析

根据 copulas 函数的定义首先应该根据两种收益率各自的边际分布把收益率数据按其分布转化成  $[0, 1]$  区间上均匀分布的数据，然后再应用这组数据获得 copulas 函数的参数估计值。当使用对数似然函数时可以把 copulas 函数

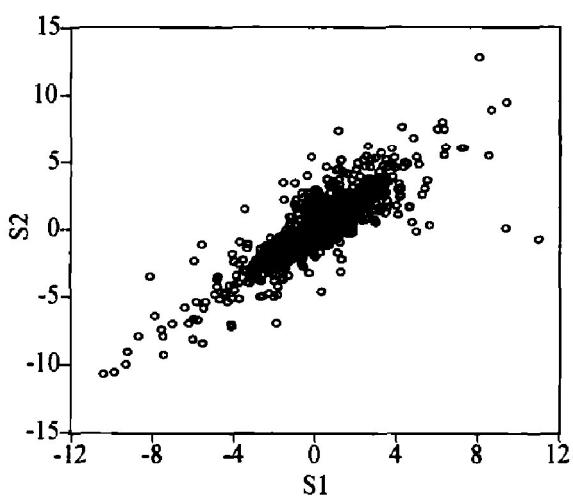


图 1 沪、深日对数收益率关系图（其中 s1 为上海收益率 s2 为深圳收益率）

的密度和边际分布函数的密度分开，从而完成先获得边际分布的参数再获得 copulas 函数的参数估计值的两步估计的目的。具体的基于二元变量的观测值组  $(x, y)$ ，其对数似然函数  $l$  形如如下的公式 (14)：

$$\begin{aligned} l(x, y, \gamma_1, \gamma_2, \delta) = & \log(c(F(x, \gamma_1), G(y, \gamma_2), \delta)) \\ & + \log f(x, \gamma_1) + g(y, \gamma_2) \end{aligned} \quad (14)$$

其中， $c$  表示 copulas 函数的密度函数； $f$ 、 $g$  分别表示两个边际分布的密度函数； $\gamma_1$ 、 $\gamma_2$  和  $\delta$  为待估参数。我们可以用极大似然估计法完成对待估参数的估计。当然，应用“三”中“copulas 函数的相关性度量”中的公式也可以完成对待估参数的估计。下面我们先用极大似然法来看我国沪、深股市指数对数收益率的 copulas 函数形式。

为了照顾到厚尾的特点，我们假设两个市场的股市指数的对数收益率服从 t 分布，这样我们分别用极大似然法估计出这两个边际分布的自由度都为 2。虽然应用 copulas 函数并不要求边际分布形式相同，但本文在都是 t 分布的前提下所得两组对数收益率所服从分布的自由度都为 2，这也与图 1 所示相符。根据边际分布可以把我们的收益率数据转化为  $[0, 1]$  区间上的均匀分布的值，如图 2 所示。

图 2 给出了两个市场的对数收益率依边际分布转化为  $[0, 1]$  区间上均匀分布的样本点的关系，横、纵轴分别表示沪、深日收益率变化的结果。通过该图可以观察到边际分布存在共同变化的趋势，并且可以看出在极值附近这种趋势更为明显。比较而言，在图 1 中我们并没有看出这么强烈的尾部相关性，但是这并不说明二者的结论相悖：这部分是由于前者相当于比较了两个市场同等

水平下的变化趋势，而后者却是考察在相同的分位数水平下的共同的变化。用公式表示前者考察的是 $\lim_{u \rightarrow \infty} p(Y > u | X > u)$ ，而后者却是 $\lim_{\alpha \rightarrow 0^+} P(Y > G^{-1}(1 - \alpha) | (X > F^{-1}(1 - \alpha)))$ 。因此，图1和图2看起来的差别，也许部分是由于沪、深市场指数的变化区间并不相同导致的。这样我们认为应该使用存在上下尾相关的 copulas 函数，故本文中介绍的 Fréchet copulas 中的  $C_u$ ，t 分布的  $C_t^{p,v}$  及 Gumbel 的  $C_G^v$  可以作为被选的 copulas 函数。并且为了比较也可以把通常被假定为正态分布的 copulas 函数考虑在内。

为了使用 t 分布和正态分布的 copulas 函数，相关系数  $\rho$  有待估计。前面我们已提到由于边际分布可能不再是简单的线性关系，相关系数也不一定是线性相关系数。为了获得此相关系数的准确值可以应用 Embrechts 等 (2001) 的结论： $\hat{\rho} = \sin(\frac{\pi}{2} \hat{\tau}_k)$ 。其中， $\hat{\tau}_k$  为 Kendall 的阶相关系数 (rank correlation)  $\tau_k$  的估计值。根据两个股票市场的收益率数据我们获得  $\hat{\tau}_k = 0.679$  这也说明了两个市场的收益率存在一定程度的正相关性，进一步可得  $\hat{\rho} = 0.876$ 。

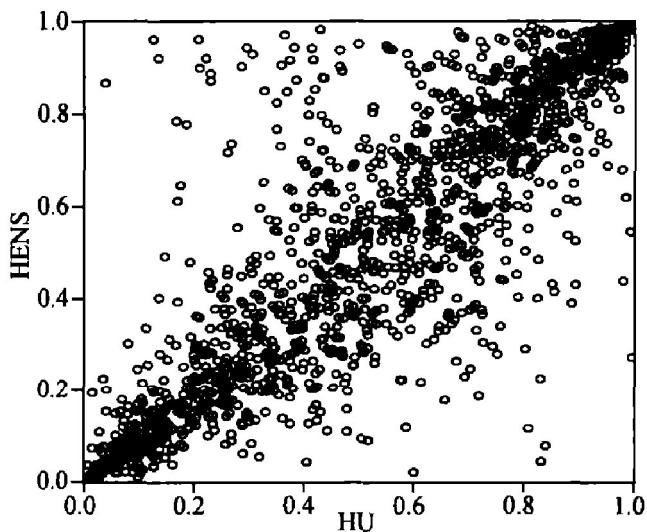


图2 天收益率数据变换为均匀分布的数据情况

对于我们考察的几个 copulas 函数还剩下  $C_t^{p,v}$  和  $C_G^v$  有待估参数。对于  $C_t^{p,v}$  我们对每个  $v \geq 2$  的整数值，求使对数似然函数达到最大的  $v$ ，如此得到  $v$  的估计值为 2，从而可以获得该 copulas 函数的上下尾相关系数为 0.72。对于  $C_G^v$  我们也应用极大似然法获得参数的估计值，并同时得到其上尾相关系数具体结果，如表1所示。

表 1 使用天数据获得 copulas 函数的极大似然估计值

	Gaussian	t-Student (3)	Gumbel	Fréchet ( $C_u$ )
$\rho$	0.876	0.876	—	—
$\theta$	—	—	—	—
$\delta$	—	—	48.6	—
$\lambda_L$	—	0.72	—	1
$\lambda_u$	—	0.72	0.9846	1

关注表 1 在边际分布都是 t 分布的假设下，得到关于正态 copulas 函数和二元 t 分布的 copulas 函数的相关系数都是 0.876，由于该相关性是稳定的我们认为在我国的沪、深市场上确实有很强的正相关性。另外，由 t 分布的 copulas 函数得到这两个边际分布有尾部相关性，同样应用 Gumbel 的 copulas 函数也可得到上尾相关性并且这个相关性较强，而如选择 Fréchet 的 copulas 函数则被考察的变量的尾部相关性就完全为 1 了，这几个 copulas 函数都在一定程度上体现了图 2 中所示的两个边际分布在极值部分的较强的相关性，但是正态的 copulas 函数却不能够有效的解释这一点，基于此我们也有理由认为正态分布的假设并不很适合对厚尾事件的研究。根据前面的介绍 Gumbel 的 copulas 函数的参数  $\delta$  值的大小能够体现被考察变量之间的相关性强弱，我们得到的估计值为 48.6，说明沪、深市场收益率的相关性较强，这也与图 1 和图 2 所体现的情况相符。因此，通过本文的实证研究我们认为 Gumbel 的 copulas 函数和 t 分布的 copulas 函数适合在对我国沪、深市场收益率的相关性进行深入的研究中使用。

### 参考文献：

- [1] Nelsen, R. B (1998), An Introduction to Copulas, Lectures Notes in Statistics, 139, Springer Verlag, New York.
- [2] Rob W. J. van den Goorbergh, Christian Genest, Bas J. M. Werker. Bivariate option pricing using dynamic copula models, Insurance: Mathematics and Economics 37 (2005) 101-114.
- [3] Roger B. Nelsen. Some properties of Schur-constant survival models and their copulas.
- [4] Charles N. Haas. On modeling correlated random variables in risk assessment. Risk Analysis, Vol. 19, No. 6, 1999.

- [5] Roger B. Nelsen. Dependence modeling with Archimedean copulas.
- [6] Roger B. Nelsen. Properties and applications of copulas: A brief survey.
- [7] Beatriz Vaz de Mendes, Rafael Martins de Souza. Measuring financial risks with copulas, Internal Review of Financial Analysis, 13 (2004), 27-45.
- [8] Fisher, N. I. Copulas. In: Encyclopedia of Statistical Sciences, Update Vol. 1 1997, 159-163. John Wiley Sons, New York.
- [9] Schweizer, B., & Sklar, A. Probabilistic metric spaces, 1983. New York: North-Holland.
- [10] Embrechts, P., Lindskog, F., & McNeil, A. (2001). Modeling dependence with copulas and applications to risk management. Zurich: Department of Mathematik, ETH Zentrum, CH 8092.
- [11] 张尧庭. 连接函数(Copula)技术与金融风险分析. 统计研究, 2002, 4.
- [12] 张尧庭. 我们应该选用什么样的相关性指标. 统计研究, 2002, 9.
- [13] 史道济, 关静. 沪深股市风险的相关性分析. 统计研究, 2003, 10.
- [14] 韦艳华, 张世英. 金融市场的相关性分析——Copula-GARCH模型及其应用. 系统工程, 2004, 4.
- [15] 韩明. Copula——一个新的计量经济工具. 统计与信息论坛, 2004, 5 (19).